KU LEUVEN

European Social Survey

Monitoring the evolution of the fieldwork power: illustration based on the seventh round of the European Social Survey.

Caroline Vandenplas, Geert Loosveldt and Koen Beullens

Comparative Survey Design and Implementation workshop, March 16-18, Mannheim

Fieldwork monitoring

- To monitor the fieldwork, follow-up on the evolution of:
  - Key performance indicators (Jans, Sirgis and Morgan, 2013):
    - effort metrics ←number of contact attempts, nbr of active interviewers
    - productivity metrics, ← number of completed interviews
    - survey output← response rate
  - 'Phase capacity' (Groves and Heeringa, 2006)

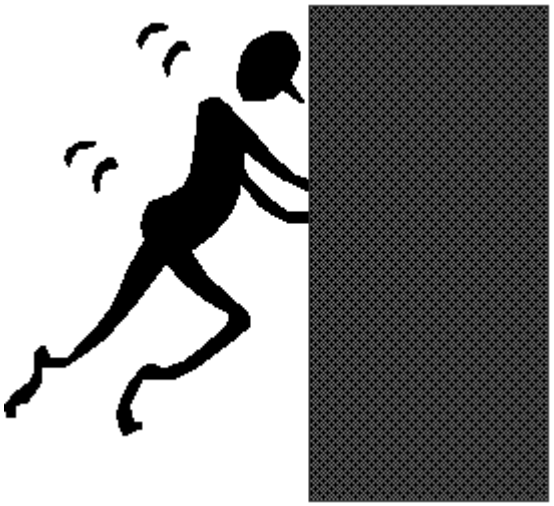
Benchmark or boundaries for monitored indicators

- To follow up the evolution of the indicators:
  - A benchmark or boundaries are needed:
    - number of contact attempts← planned, budgeted for
    - number of completed interviews ←? expectations
    - response rate ← given threshold
  - Phase capacity← look at the variations…
- Boundaries or benchmark are based on knowledge/information

Benchmark or boundaries for monitored indicators

- A benchmark can be developed based:
  - General knowledge of stakeholders or technicalities
  - Information on
    - Sampling units: based on the sampling frame (gender, locality, age) or collected during the fieldwork (current status)
    - The fieldwork in general: based on previous rounds, similar surveys, same surveys in similar countries or previous ‘phase’ of the same fieldwork

Idea: instead of monitoring cumulative indicator, monitoring of the indicator per time unit



Final number of completed interviews

Work= Power X Time

(Mean) Weekly number of completed interviews

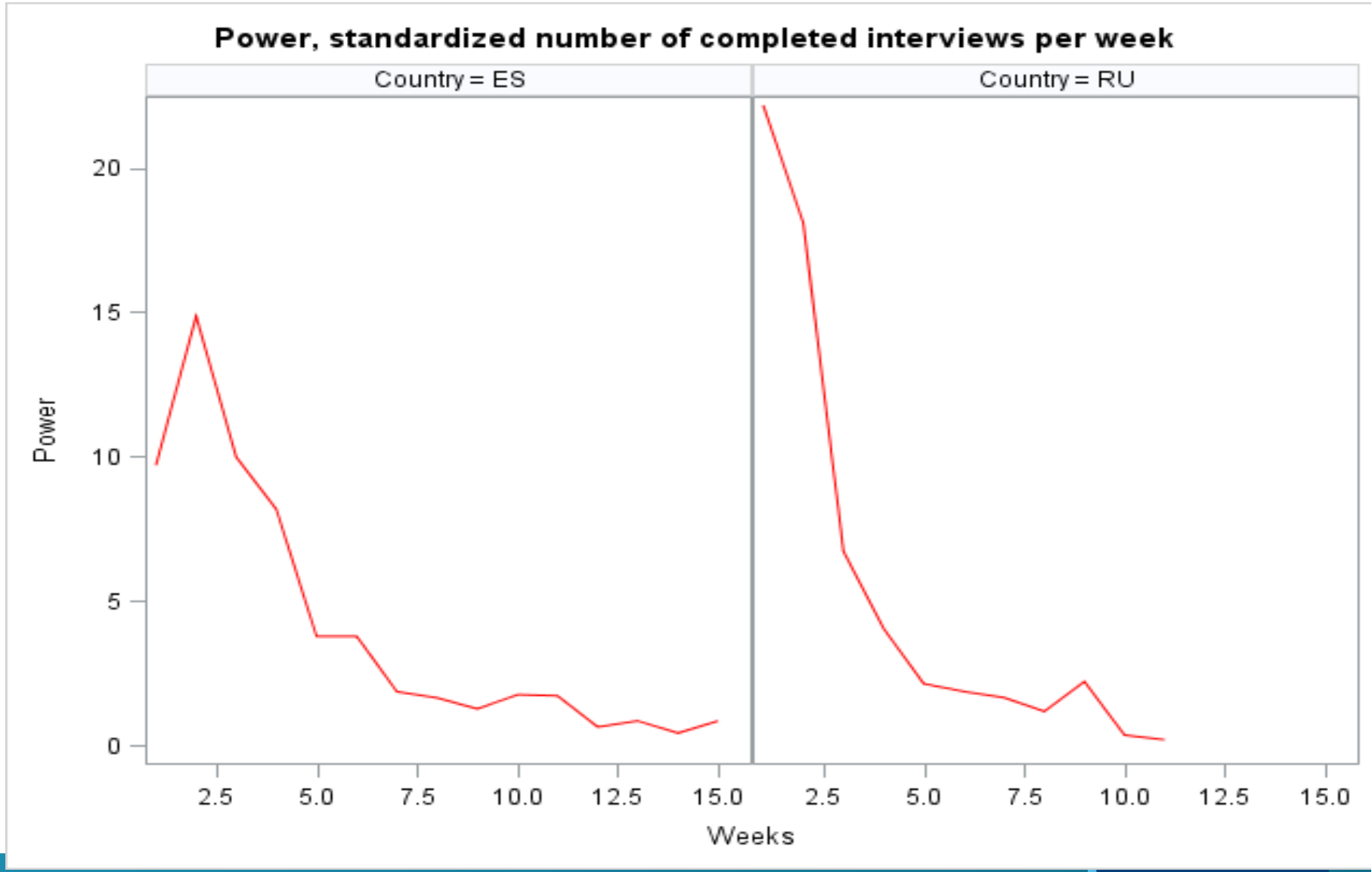
Fieldwork period (weeks)

KU LEUVEN

The fieldwork power as a productivity metric

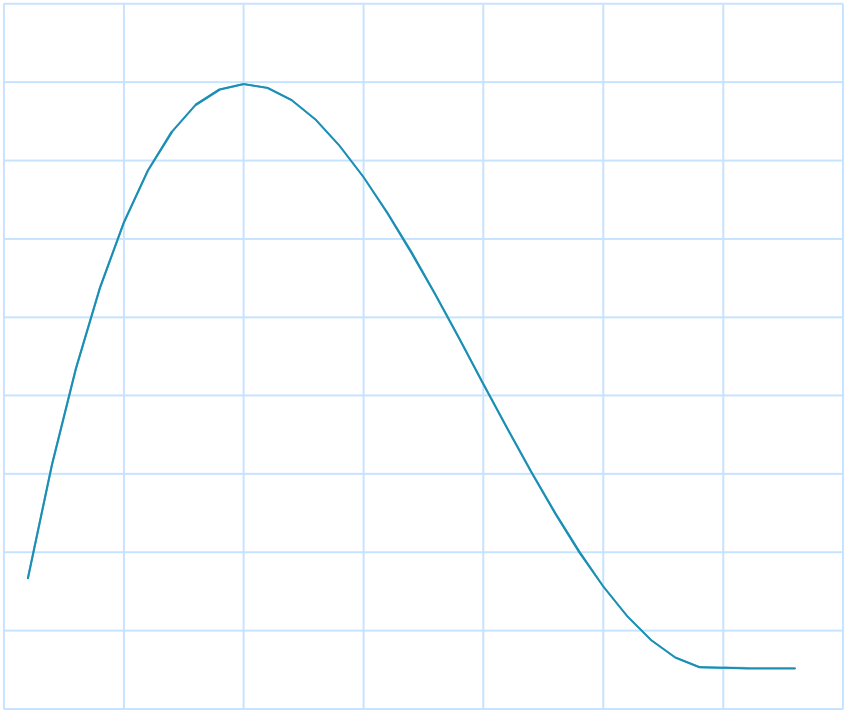
- Yield of the fieldwork *per time unit:*
  - The fieldwork power can be defined in various ways:
    - The number of completed interviews per time unit
    - The number of contacts established per time unit
    - The ratio of number of completed interviews and number of contact attempts per time unit
    - The ratio of number of completed interviews and number of refusals per time unit
  - The time unit can be defined in different ways:
    - Frequently enough to catch the dynamic
    - Spaced enough to have the time to gather information and avoid irrelevant fluctuations
    - For the ESS, a face-to-face survey, we will work with weeks

Modeling the fieldwork power to create a benchmark



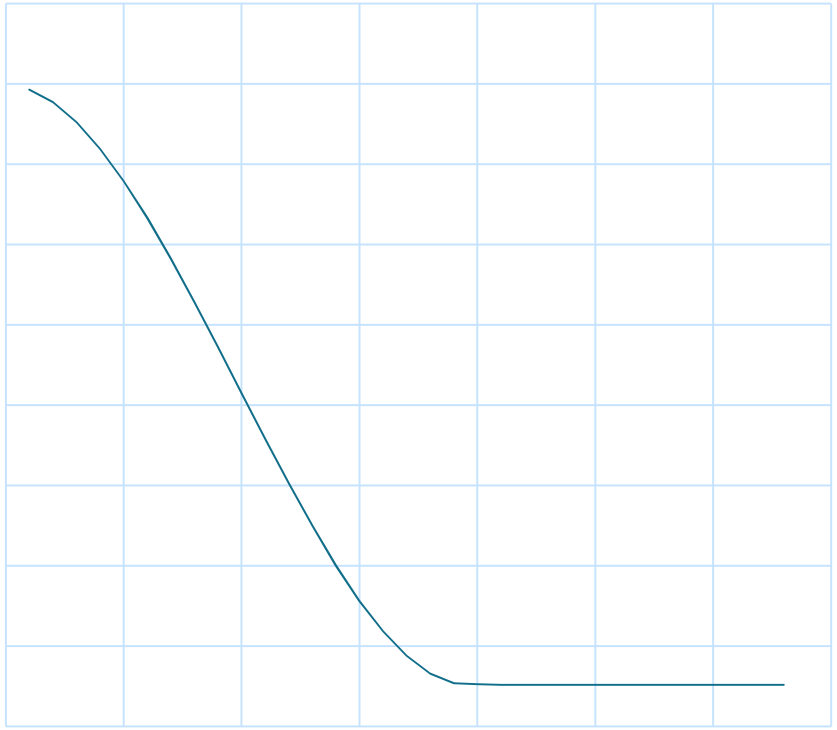
General shape of the fieldwork power

Spain



weeks

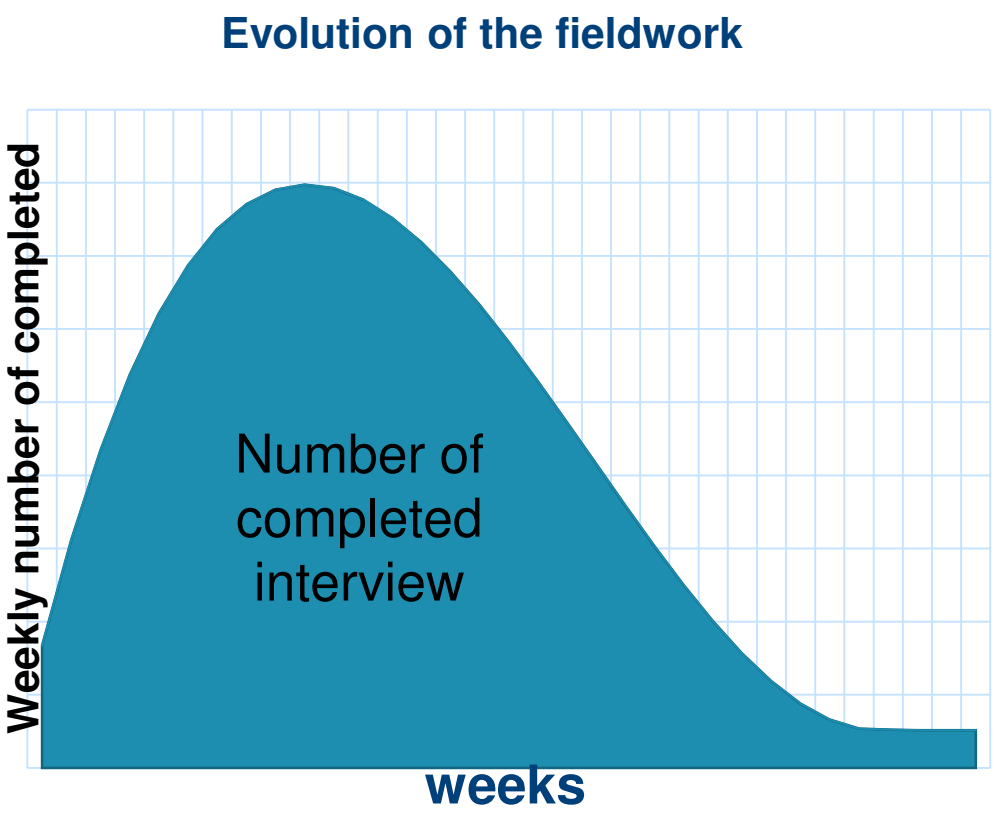
Russia



weeks

KU LEUVEN

Time dependent Power…



Standardize the number of sampled units to 100 for cross-survey comparison

Final nbr of completed interviews=

$$\sum_{Fieldworkweeks} Nbr\ of\ completed\ interviews\ in\ week\ w$$

Model the evolution of the fieldwork power measurements

- We model the power of surveys in the European Social Survey. There are in total 149 surveys (country-round combinations) in the first six rounds
- For each fieldwork week of each survey, we have one measurement of ‘power’
- Four important characteristics in the evolution of the fieldwork power:
  - The starting power
  - The starting increase or decrease in power (speed)
  - The starting decrease in speed
  - The start of the tail

Multi-level models with repeated measurements

- The macro-level are ESS surveys: combination of rounds and countries participating in that round
- The repeated measurements are the weekly fieldwork power as specified for each considered ESS survey
- The model:

$$P(s,w) = \beta_0(s) + \beta_1(s)w + \beta_2(s)w^2 + \beta_3 w^3 + \varepsilon_{s,w},$$

$$\beta_0(s) = \gamma_{00} + u_{0s},$$

$$\beta_1(s) = \gamma_{10} + u_{1s},$$

$$\beta_2(s) = \gamma_{20} + u_{2s},$$

$$\beta_3 = \gamma_{30},$$

Three benchmark levels

- ESS curve: 149 ESS surveys from the first six rounds
- ‘Similar surveys’ curve - ESS surveys’ with following characteristics:
  - Individual vs non-individual sampling frame
  - Percentage of refusal conversion
  - Response rate
- Previous rounds benchmark :Surveys from previous ESS rounds in the same country
- Why three benchmarks? Precision vs accuracy, different countries may have different information

Constructing the benchmark curves

- For each level, enter the corresponding surveys into the model:

$$P(s,w) = \beta_0(s) + \beta_1(s)w + \beta_2(s)w^2 + \beta_3 w^3 + \varepsilon_{s,w},$$

$$\beta_0(s) = \gamma_{00} + u_{0s},$$

$$\beta_1(s) = \gamma_{10} + u_{1s},$$

$$\beta_2(s) = \gamma_{20} + u_{2s},$$

$$\beta_3 = \gamma_{30},$$

- Use the parameter estimates of $\gamma_{00}, \gamma_{10}, \gamma_{20}, \gamma_{30}$ to construct the benchmark curve

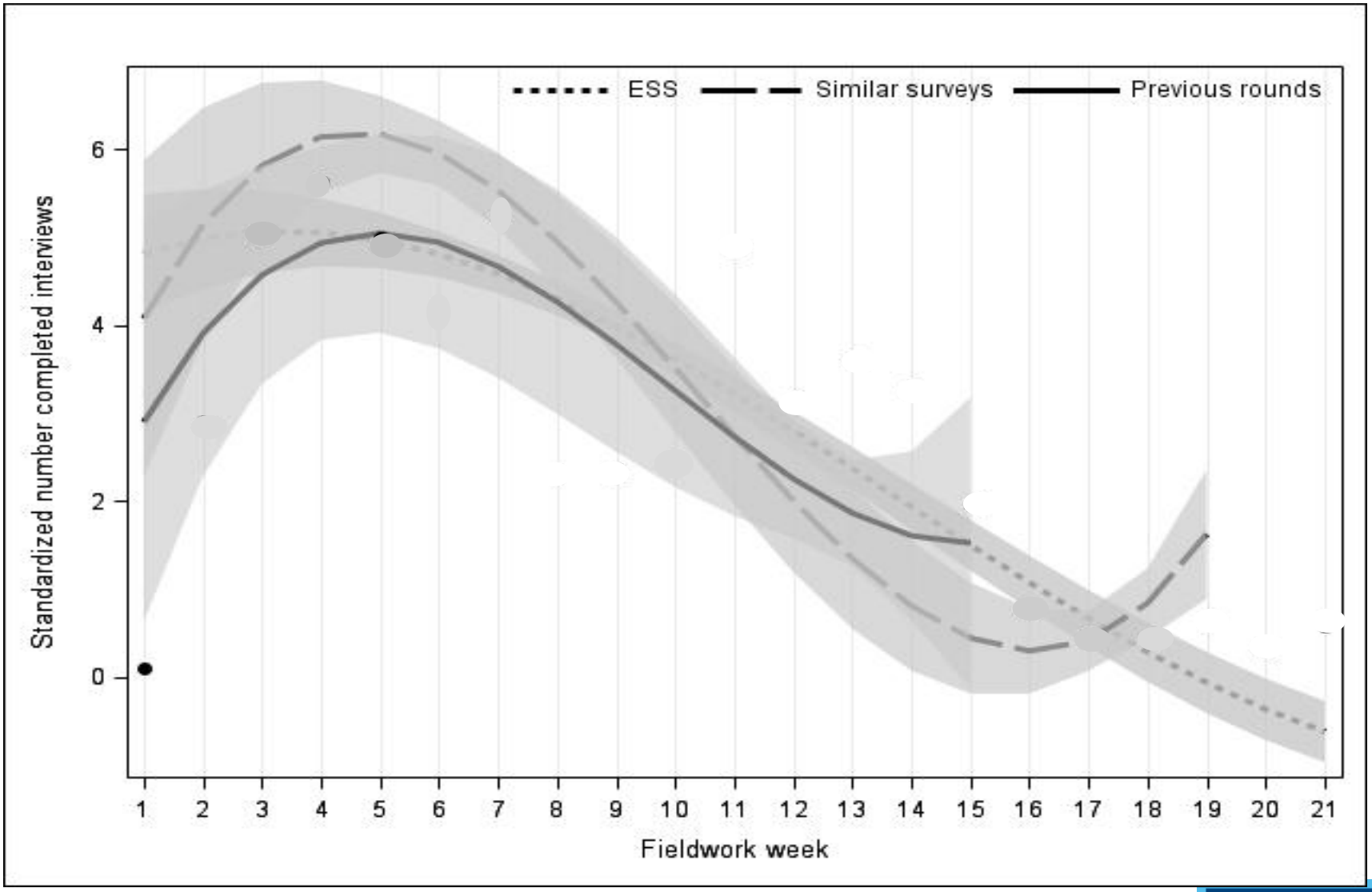
$$\gamma_{00} + \gamma_{10}w + \gamma_{20}w^2 + \gamma_{30}w^3$$

And the corresponding confidence band.

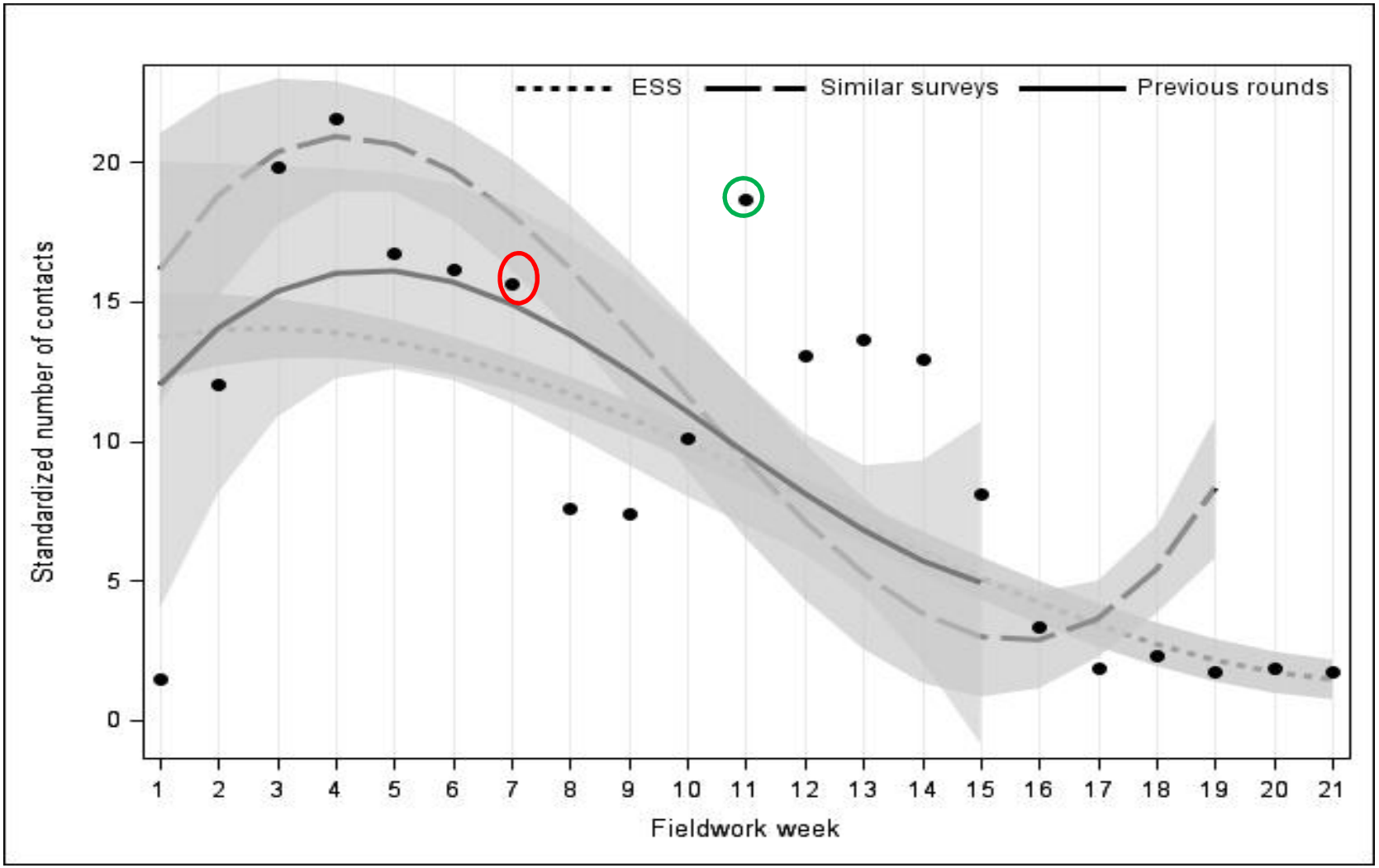
Flagging rules

- Immediate action should be taken if the fieldwork power (any of the four specifications):
  - is below the confidence band of the benchmark in two subsequent weeks;
  - is below the benchmark for three weeks in a row;
  - or, reduces for three weeks in a row.

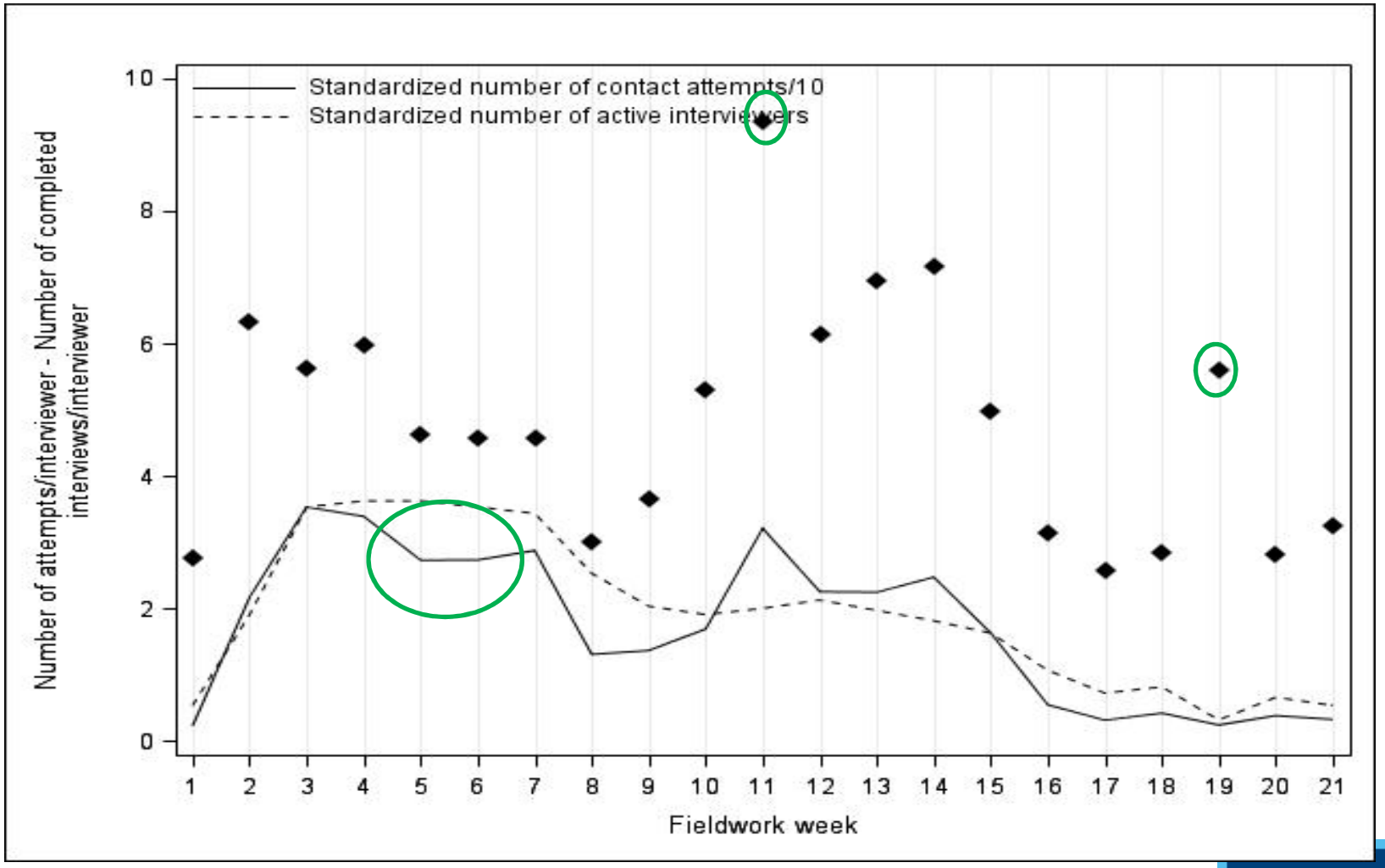
Belgium in round 7: completed interviews



BE R7: contacts



BE R7: effort metrics



KU LEUVEN

Data quality indicator

In parallel to the fieldwork power, we monitor data quality indicators:

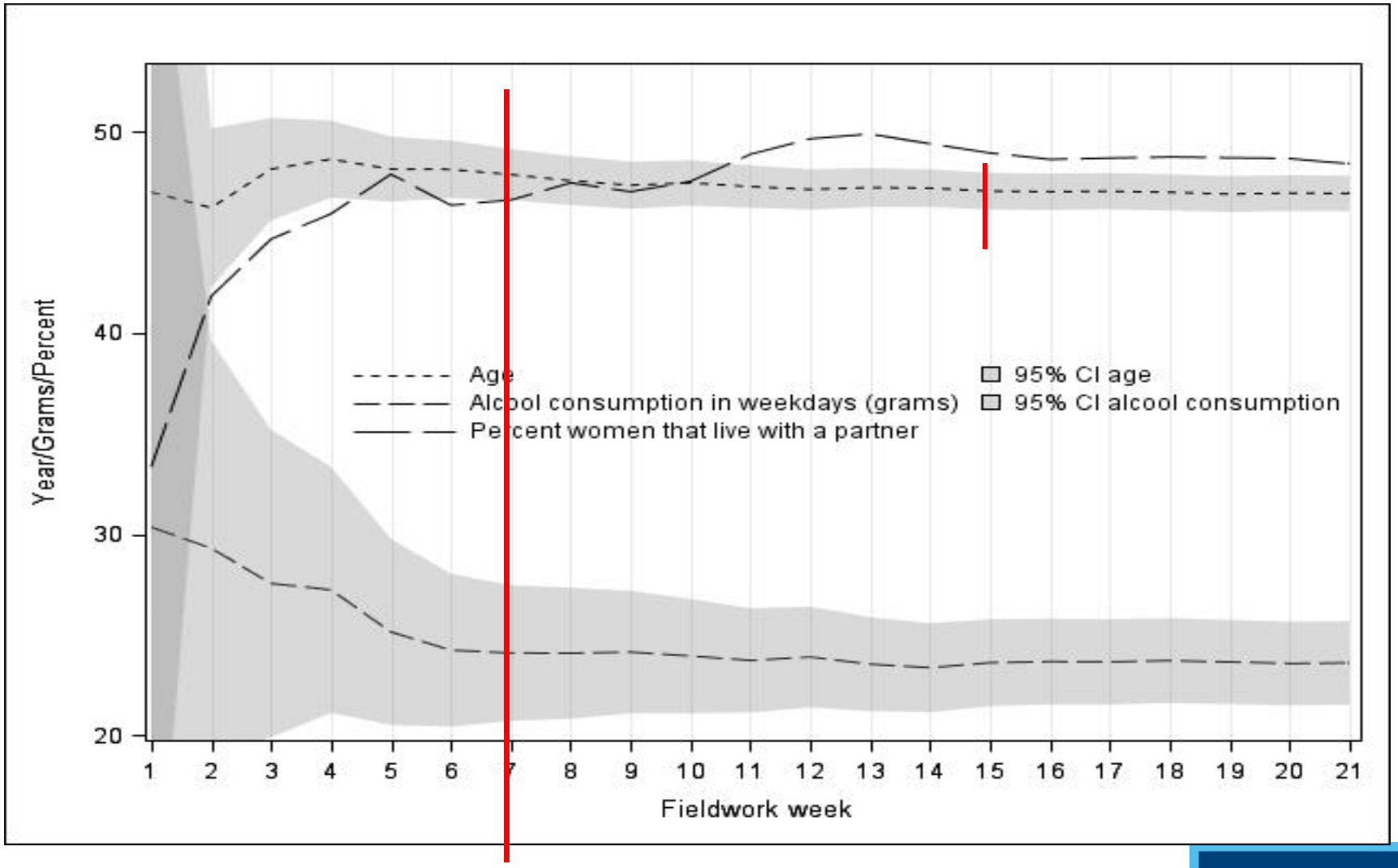
- Age and it's SE
- Alcohol consumption (rotating module) and it's SE
- Percentage of woman amongst respondent with a partner

Flagging rules

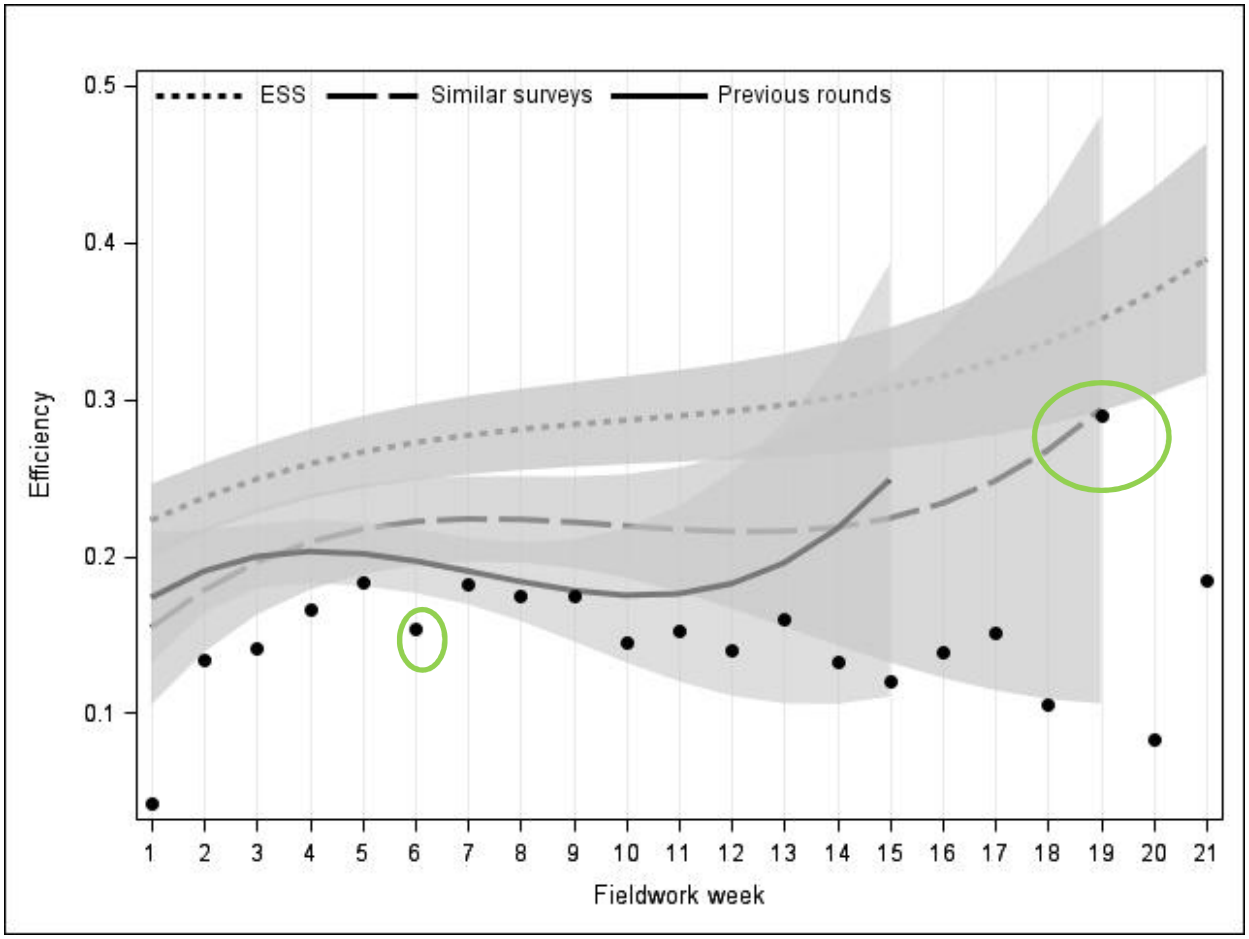
The fieldwork has reached is phase capacity if;

- The sampling error of the considered variable is lower than $SE_{pre} = \sigma/\sqrt{1500}$ for two weeks in a row,σ is calculated based
  - on the standard deviation estimates of other sources as for instance the previous round (age)
  - On the standard deviation estimates based on the data obtained so far (alcohol consumption)
- the absolute difference in the estimate of a week from that of the previous one is lower than SE_{pre} for two weeks in a row.

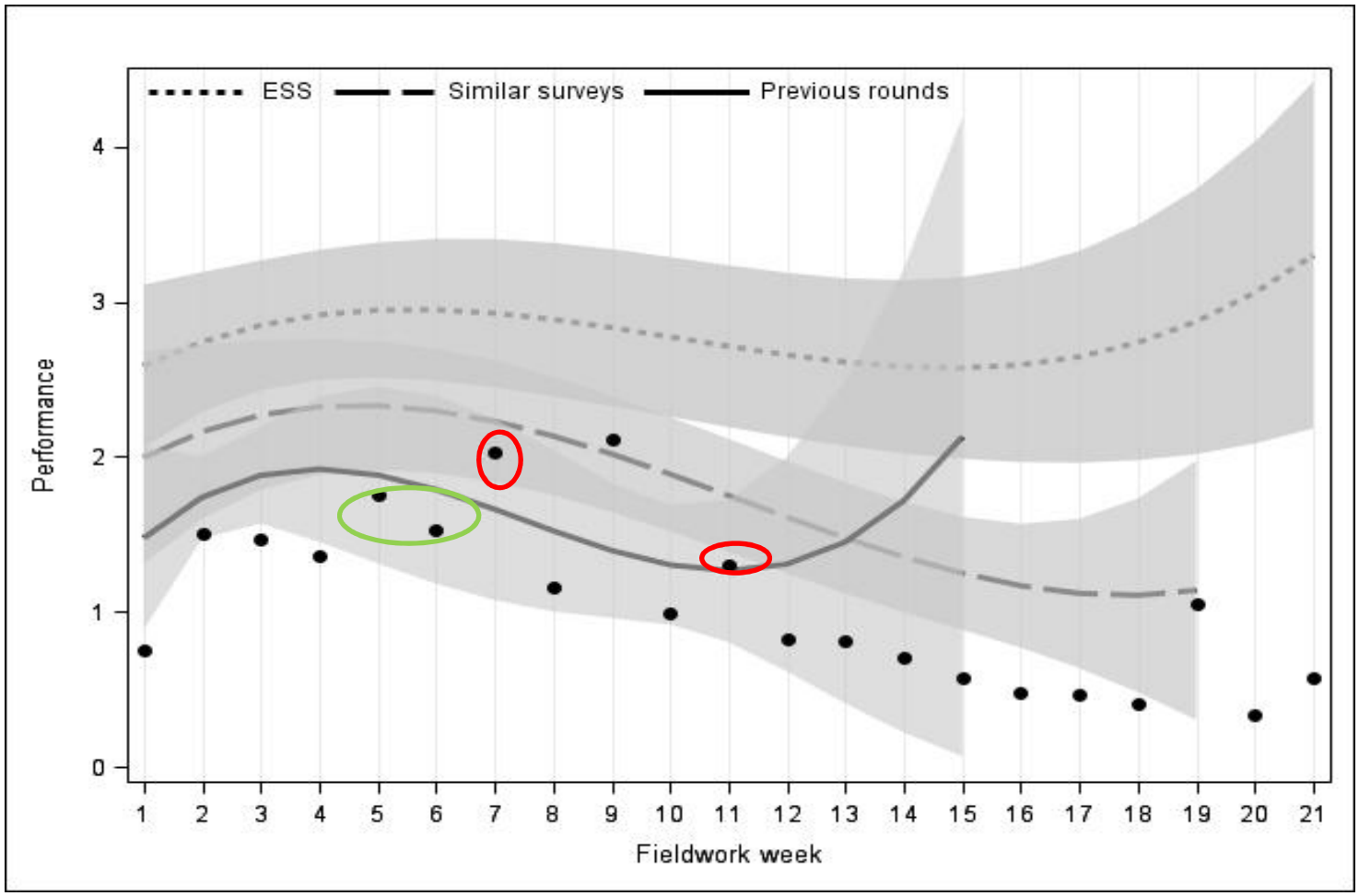
BE R7: data quality metric



BE R7: Efficiency (contacts/attempts)

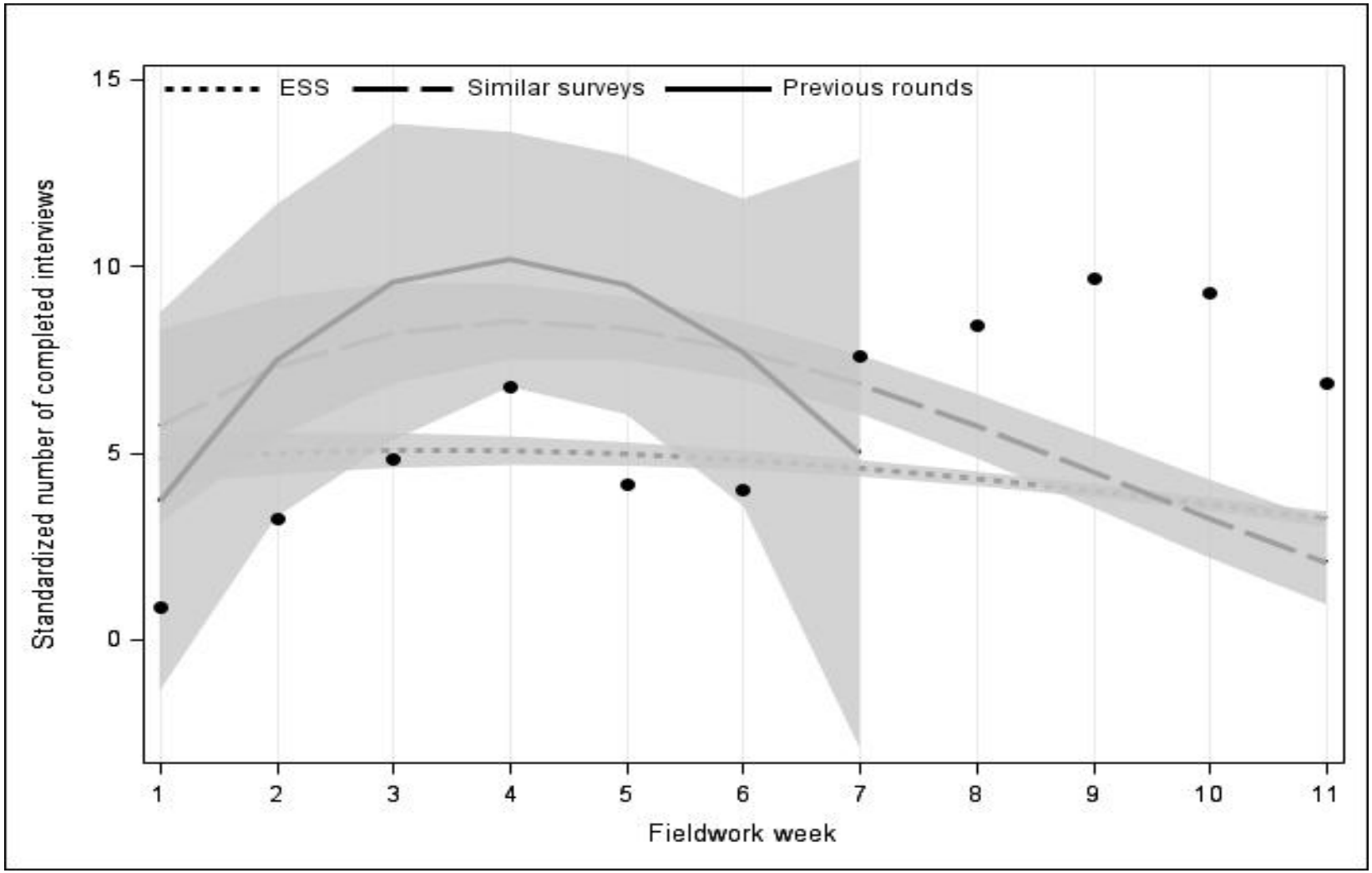


BE R7: Performance(completed/refusals)

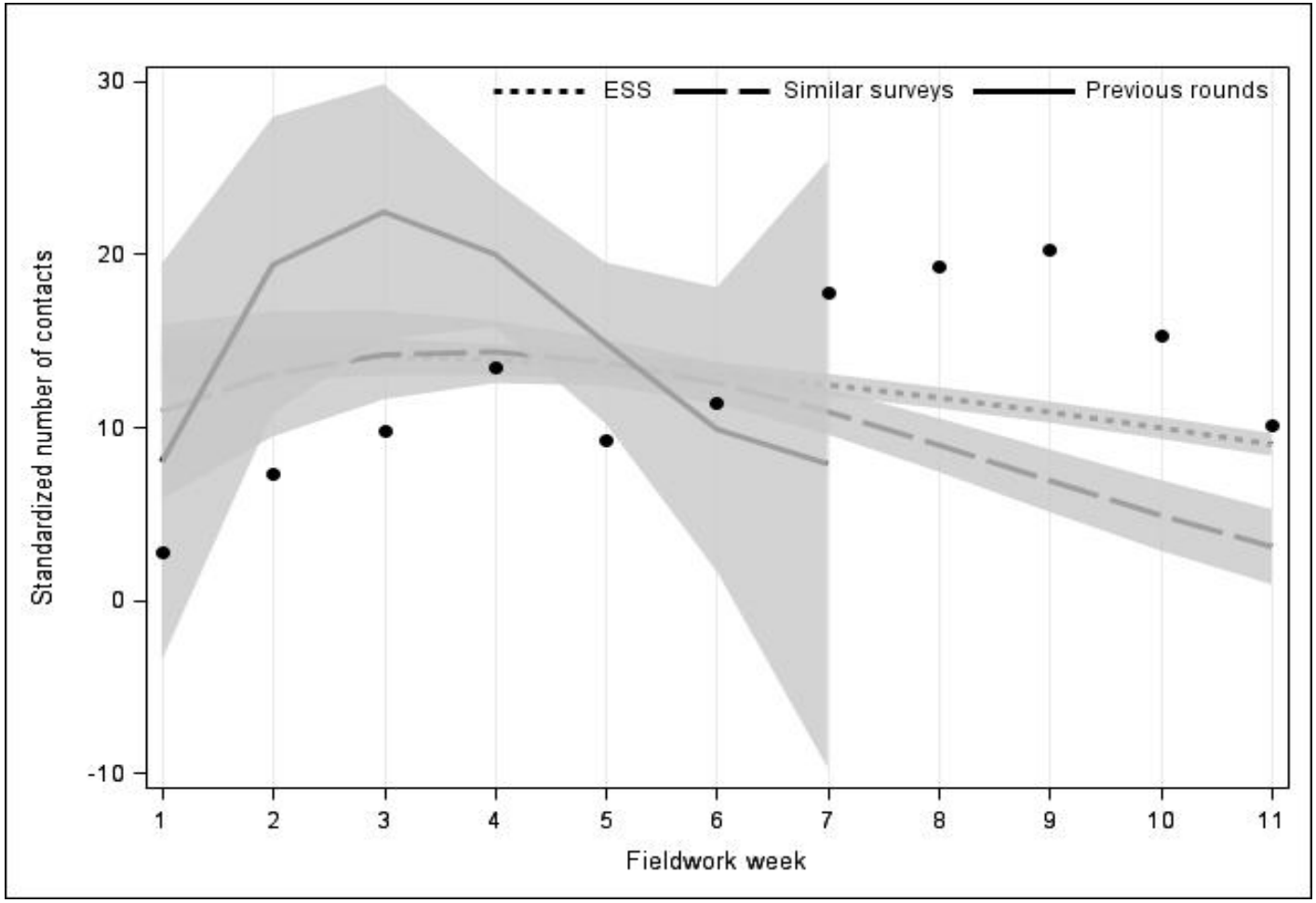


KU LEUVEN

The Czech Republic Round 7: Completed interviews

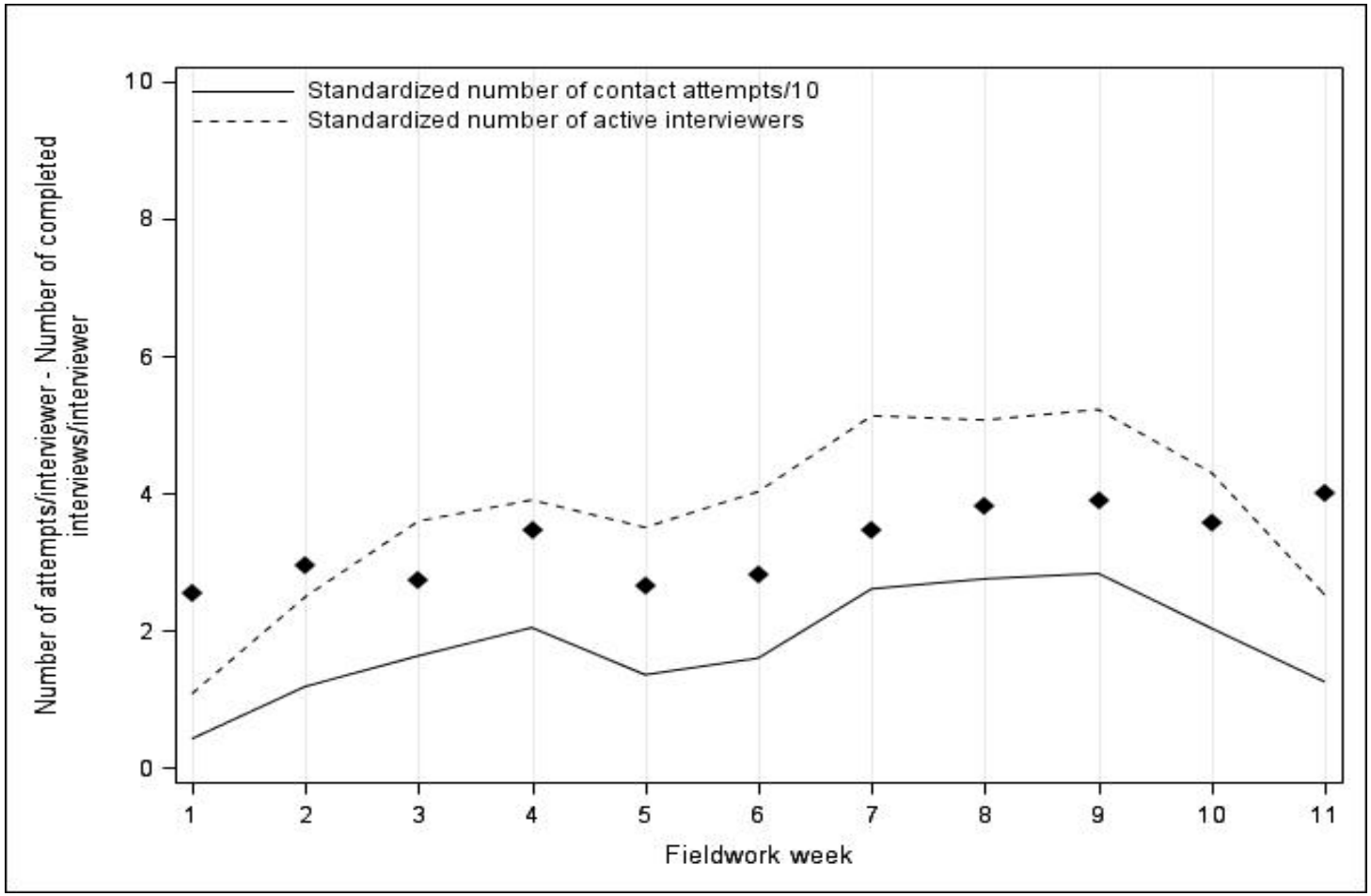


CZ R7: contacts

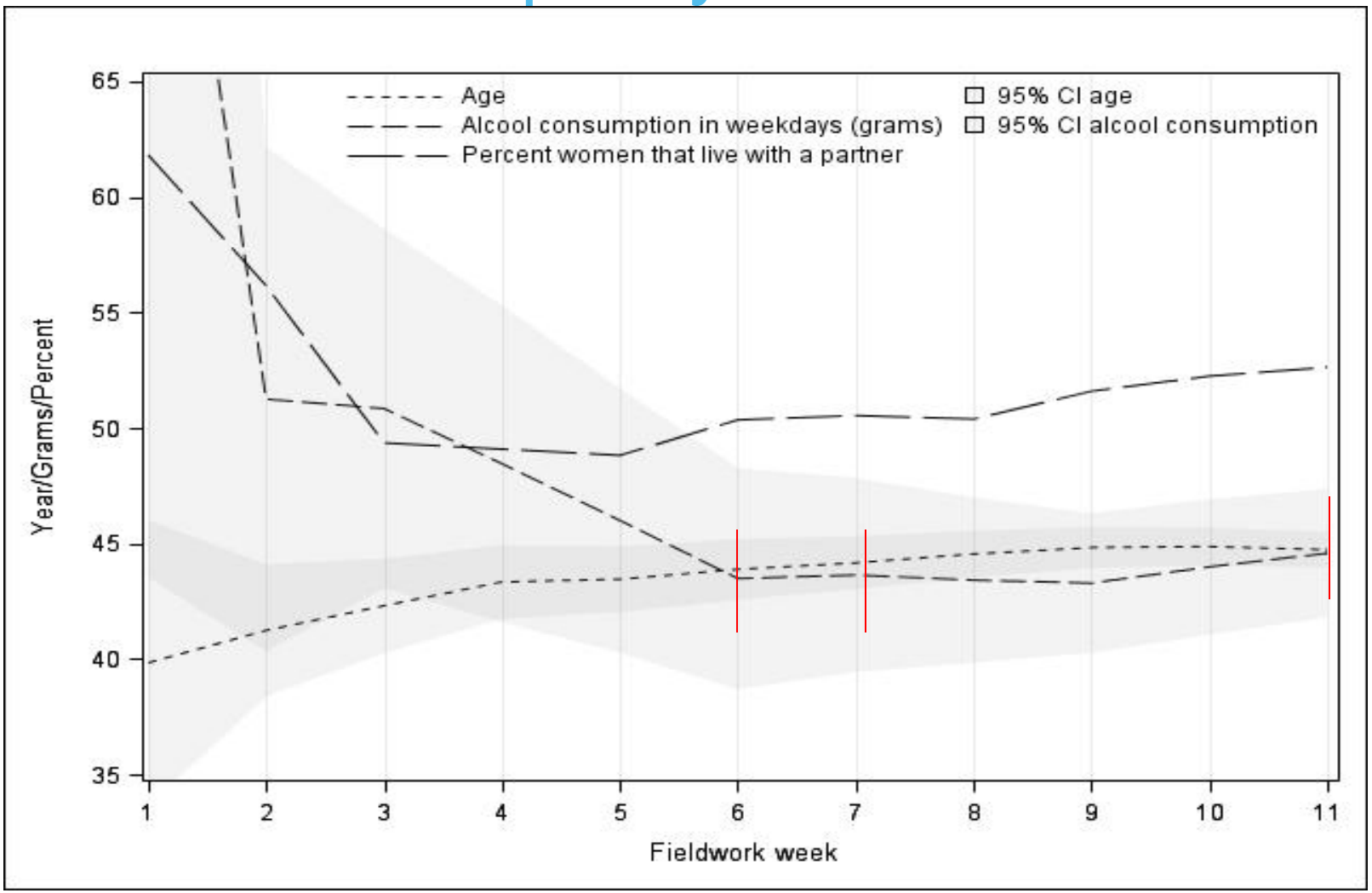


KU LEUVEN

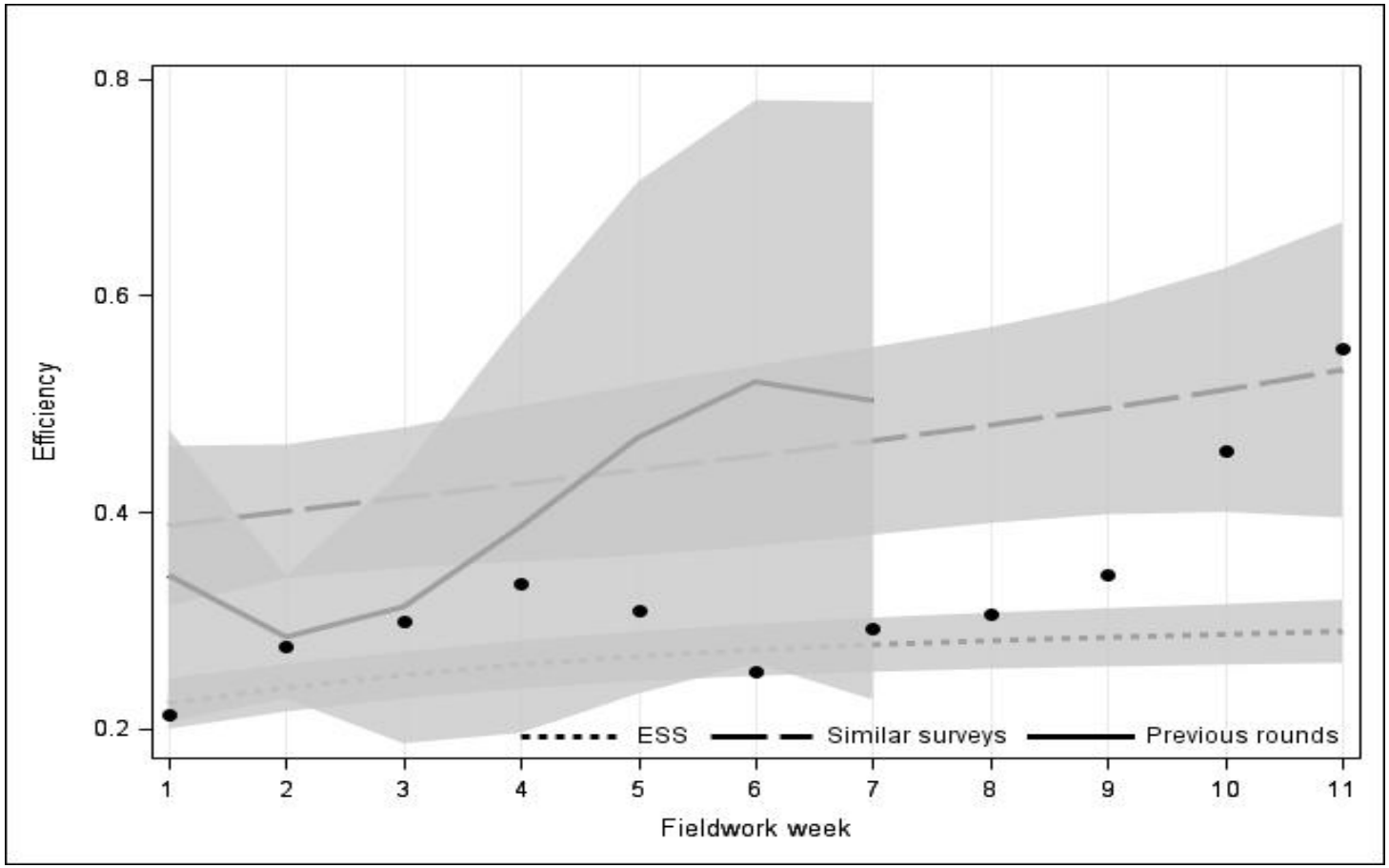
CZ R7: effort metric



CZ R7: data quality



CZ R7: efficiency(completed/attempted)



KU LEUVEN

Conclusion from monitoring the fieldwork power

- Completed and contacts:
  - Clear deviating pattern compare to the benchmark, lower in the first six weeks and higher later (weeks 8,9,10), no tail
  - Efficiency highest at the end of the fieldwork

- Data quality:
  - Sampling error threshold only reached in the last week for age
  - The percentage of women with a partner increase above 50% after week 8

Overall conclusions

- The benchmarks created with the multi-level models help detecting deviating patterns during the fieldwork and as post-survey evaluation
- Further work:
  - Feasibility of ‘live’ monitoring in ESS
  - Apply to other survey designs
  - Other definition of fieldwork power (new contacts)
  - Correlation between data quality and fieldwork power
  - Development of other type of metrics

Interventions

- The interventions when a week is flagged should be planned and budgeted before the fieldwork
- But what can we do?
  - Cause of the flag?
    - To low effort (not enough interviewer or too low effort from the interviewer part)→ re-called/retrained interviewer, redistribution of (new) adresses, giving feedback to interviewer on their performance compared to other interviewers
    - To low efficiency performance→ Incentive?, redistribution of hard cases to the best inteviewer, marketing?

Caroline.vandenplas@kuleuven.be

KU LEUVEN